University of British Columbia & Simon Fraser University – The Bricolage

William Chao, Daniel Ha, Kevin Ho, Linda Kaastra, Minjung Kim, Andrew Wade*, Faculty Sponsor: Brian Fisher

Simon Fraser University, SIAT and University of British Columbia, MAGIC

ABSTRACT

This abstract presents a *bricolage* approach to the 2007 VAST contest. The analytical process we used is presented across four stages of sensemaking. Several tools were used throughout our approach, and we present their strengths and weaknesses for specific aspects of the analytical process. In addition, we review the details of both individual and collaborative techniques for solving visual analytics problems.

CR Categories and Subject Descriptors: H.3.3 [Information Search and Retrieval]: Information filtering, Relevance feedback, Selection process; H.4.1 [Office Automation]: Groupware; H.5.3 [Group and Organization Interfaces]: Collaborative computing, Computer-supported cooperative work

Additional Keywords: bricolage, collaboration, theory building, meta-analysis, evaluation

1 INTRODUCTION

Our group solved the VAST 2007 contest problem with a *bricolage* approach, using a variety of pre-existing tools in combination with one another throughout the different stages of the analytic process. These tools were not pre-selected for their features, but rather were found and used as needed. Much of the initial analysis was performed individually; however collaboration between analysts was crucial to later stages of analysis and ultimately to the generation of a unified hypothesis.

2 SOLUTION TO THE VAST 2007 CONTEST PROBLEM

From the dataset, we determined the key players involved in illegal and terrorist activities, a timeline of events and locations important to the plot. Cesar Gil, a biologist and animal rights activist, was believed to be the main culprit. In order to thwart a new fad of chinchillas as pets, Gil orchestrated a monkeypox outbreak that spread to humans in the LA area from infected chinchillas. Gil's hope was that people would begin to fear chinchillas as disease-carriers and the fad would quickly end. The distribution of infected chinchillas can be traced to a company called Global Ways. Global Ways is an import/export company that is also involved in the illegal smuggling of drugs and animals from South America and Africa. Abu Hassan, owner of a circus in Africa, was implicated in animal smuggling and has exclusive ties to Global Ways. Navarro Mercurio, operations manager for the Miami office of Global Ways is involved in illegal smuggling of cocaine and animals. The infected chinchillas are brought into the country by a known chinchilla poacher, Rosalind Baptista. Baptista met with Mercurio in New Orleans to arrange for the chinchillas to enter the country. Global Ways CEO Madhi Kim is also involved with illegal smuggling activities. r'Bear, a rapper who owns the Shravaana ranch outside the LA Area, uses Madhi Kim and Global Ways to import animals. As a result, r'Bear receives some of the infected chinchillas brought into the country by Baptista and Mercurio. r'Bear becomes infected with monkeypox and is hospitalized shortly before the outbreak in the LA area. The rest of the infected chinchillas are delivered to Gil in LA, who distributes them through his chinchilla farm, Gil Breeders. Another person of interest is Luella Vedric, a New York socialite. Due to her associations with both Madhi Kim and organizations like SPOMA, we recommended she be investigated.

*e-mail: andrewwade@ubcviscog.com

3 DESCRIPTION OF ANALYTICAL PROCESS

Our analytical process can be described in terms of four distinct but overlapping stages: information generation, schematization, argumentation and schema shifting, and decision-making [1]. Our problem solving approach gradually shifted from independent to collaborative work as we progressed through the stages. The shift was deliberate and the tools chosen along the way reflect this change.

3.1 Information Generation

3.1.1 Information Discovery and Foraging

This part of the Information Generation stage was characterized by finding categories for the information and coding for them in the data, including names, places and events. Initially, analysts were encouraged to work independently in order to mitigate the potential for groupthink: a faulty, conforming style of group analysis that can lead to poor decision making [2]. The goal was for analysts to form independent categorizations and hypotheses. The tools used in this initial stage were largely entity extraction and search tools, including Windows Explorer, TextSTAT and Search Utility. An additional technique used by some analysts was automatic labeling of word forms. This was accomplished using the Stanford POSTagger, which parses a text file and appends a tag to every word present, thereby identifying its part-of-speech (e.g., noun, verb). This allowed for easier isolation of proper nouns and terms of interest. Another tool used for early coding and extraction of entities was the commercial tool ATLAS.ti.

In order to move towards a collaborative answer, analysts began to share their findings. ATLAS.ti provided a useful network diagram view of relations between entities that facilitated communication of ideas between analysts. Other visual representations of individual findings were created with FreeMind, which allowed for an intuitive depiction of many ideas branching from a single lead as a divergent node.

3.1.2 Information Pooling, Data Transformations and Integration

Tools used in this part of Information Generation represented the need to create a collective understanding of the nature of the problem. Pooling of data and analytic work took place in both physical and virtual space. Sticky notes as well as a paper timeline were used to display findings. Although they were "low tech,"

these representations were powerful in facilitating communication between different analysts, providing a space where information was exchanged through speech, pointing, gestures, and even body language. In addition, an online wiki and a shared Google Spreadsheet served as the virtual space used for pooling information and hypotheses. Using Google Spreadsheet, we could see the evolution of our analytical process: viewing group entries to the database in real time, responding to them and having spontaneous discussions emerge over particular findings.

A final tool used for communicating findings in this stage was Excel. One of the most useful data representations was a simple visualization showing the geographical locations of important characters over time.

3.2 Schematization

This stage consisted of converting pooled information into diagrams, placing emphasis on patterns and connections of interest. The majority of visualizations created in this stage were network diagrams, with links indicating a relationship between two entities. We used a tool called GraphViz to automatically generate network diagrams based on our pooled data in the Google Spreadsheet. With these diagrams, the strength of the connection was represented through color, based on our own subjective ratings.

Another tool used for generation of network diagrams was Cmap Tools. We used Cmap Tools to create a diagram of motivations for both people and organizations. These were based on facts from the data, and were used to make inferences about the probability of certain characters being involved in illegal activities.

The final tool used in this stage was a trial version of Timeline Maker. This tool allowed for the creation of an easily changeable visual representation of the temporal order of events. It also allowed for coding events in different categories, representing categories by color. This was useful when following a timeline of events for a single character, for example.

3.3 Argumentation and Schema Shifting

Agreeing upon and forming a coherent and unified hypothesis was not a simple process. In order to eliminate incorrect hypotheses, we engaged Richard J. Heuer's model for analyzing competing hypotheses [3]. We used Google Spreadsheets as a template, enabling simultaneous collaboration. We found that framing hypotheses in a manner suitable for Heuer's model was challenging given the dataset. Evidence from the data often supported multiple hypotheses, making the process of elimination very difficult. However, use of Heuer's model proved successful in eliminating speculative evidence and formulating more specific hypotheses.

3.4 Decision-Making

The final stage involved a creating a "skeleton hypothesis" that included only facts and inferences we were very confident about. This served as a template for the final submission. Many of the previous visualizations were used in deciding on the final hypothesis. Much of the decision-making near the end of the analysis happened face-to-face, and previously computer-based visualizations facilitated the consensus-making process that ultimately led to our proposed solution.

4 Conclusion

Rather than create a single tool to solve the VAST 2007 contest problem, we chose to use a diverse set of pre-existing tools addressing specific needs that arose throughout the analytical process. Though not particularly powerful when used alone, these tools combined with one another and used intelligently by analysts proved to be more than sufficient for solving the problem. Furthermore, we found that none of the tools were useful at every stage of the group sensemaking activity, which was characterized by overlapping phases of independent and collaborative work. The bricolage approach is powerful because tools are selected for features that complement analysts' requirements over time. This exercise highlights the need to better understand the human side of visual analytics: tools do not carry out the analytical process, but rather serve to facilitate it. Problem solving is a task ultimately left for human analysts and their collaborative thinking.

References

[1] D.M. Russell, M.J. Stefik, P. Pirolli, and S. K. Card. The Cost Structure of Sensemaking. Paper presented at the INTERCHI '93 Conference on Human Factors in Computing Systems, Amsterdam, The Netherlands, pp. 269-276, April 24-29, 1993.

[2] Janis, I. L. (1982). Groupthink : psychological studies of policy decisions and fiascoes (2nd ed.). Boston: Houghton Mifflin.

[3] R. J. Heuer, Psychology of Intelligence Analysis. Washington, D.C.: Central Intelligence Agency Center for the Study of Intelligence, 1999. [Online]. Available: https://www.cia.gov/library/center-for-the-study-of-intelligence/csi-publications/books-and-monographs/psychology-of-intelligence-analysis/index.html